

**Manuscript version: Author's Accepted Manuscript**

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

**Persistent WRAP URL:**

<http://wrap.warwick.ac.uk/109629>

**How to cite:**

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

© 2018, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



**Publisher's statement:**

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk).

# Top-k Selection With Pairwise Comparisons

Matthew Groves\*

*Mathematics Institute, University of Warwick, Coventry, UK*

Juergen Branke

*Warwick Business School, University of Warwick, Coventry UK*

---

## Abstract

In this work we consider active, pairwise top- $\kappa$  selection, the problem of identifying the highest quality subset of given size from a set of alternatives, based on the information collected from noisy, sequentially chosen pairwise comparisons. We adapt two well known Bayesian sequential sampling techniques, the Knowledge Gradient policy and the Optimal Computing Budget Allocation framework for the pairwise setting and compare their performance on a range of empirical tests. We demonstrate that these methods are able to match or outperform the current state of the art racing algorithm approach.

*Keywords:* Preference Learning, heuristics, simulation, subset selection

---

## 1. Introduction

Top- $\kappa$  selection is a well known problem, with applications in many different areas, including player ranking in games, selection in evolutionary algorithms, optimising search engine result relevance, and preference elicitation in decision making or other social contexts.

In the standard problem, the score or “quality” of each of  $K$  possible alternatives is modelled by the expectation of a real-valued random variable, a statistic estimated through repeated sampling. The setting of the problem can be static or active: a set of sampling results may be provided a priori, or the ranker may be allowed to sequentially select which alternatives to sample as the algorithm progresses. The aim is to efficiently and accurately select the best subset of given size  $\kappa$  from the set of alternatives. In the active setting, the total number of samples available is generally restricted, giving rise to an optimisation problem, with the objective of devising sampling procedures to maximise

---

\*Corresponding author

*Email addresses:* `M.J.Groves@warwick.ac.uk` (Matthew Groves),  
`Juergen.Branke@wbs.ac.uk` (Juergen Branke)

the probability of correctly selecting the highest scoring alternatives or subset of alternatives within the sampling budget constraint.

However, in many real world applications it can be difficult or impractical to directly estimate an alternative’s quality through sampling. Instead, it may only be possible to obtain pairwise information, either in the form of a numerical value, or as a binary preference, expressing the result of a comparison between two items. For a motivating example, consider the ranking of two football teams; it is unclear how one might accurately assess the strength of each team in isolation, but by playing the teams against each other and recording the result, we obtain pairwise information that can be translated into a ranking. Thus, we consider an adaptation of the standard top- $\kappa$  selection problem that restricts the sampling process to allow only pairwise comparisons between alternatives. Here, rather than modelling the score of an alternative as a random variable that can be sampled directly, we instead treat the outcome of each possible pairwise comparison between alternatives as a random variable. The score of an alternative is then considered to be the sum of the expectation of the  $K - 1$  R.V.’s for the pairwise comparisons with all other alternatives.

This sampling restriction increases the complexity of the problem. In general, the number of individual samples required to obtain a single measurement of the score of all possible alternatives increase from  $K$  to  $\frac{K(K-1)}{2}$ . In addition, the outcomes of the pairwise comparisons need not be transitive. For example, in the context of game players, differences in playing styles and counter strategies might create cycles in pairwise performance (A beats B, B beats C, C beats A). The information gained from a particular pairwise comparison against a particular opponent thereby only relates to part of an alternative’s overall quality, leading to additional complications when attempting to optimise the sampling process.

Our contributions are as follows:

1. We propose two novel active sample allocation methods for top- $\kappa$  subset selection, adapted for pairwise sampling problems from the well-known Optimal Computing Budget Allocation and Knowledge Gradient frameworks.
2. We prove that the proposed methods are asymptotically optimal under certain conditions
3. We empirically investigate the performance of the proposed sampling methods in various settings, and compare against current state-of-the-art pairwise subset selection methods.

In the following section, we describe the problem in more detail. Section 3 gives a brief overview of current approaches to both static and active pairwise subset selection problems, before an overview of two myopic sampling strategies that we adapt to the pairwise problem in Section 4. We then apply each of these strategies to various problem settings, with the results of empirical testing discussed in Section 5. We conclude in Section 6.

## 2. Problem Definition

The problem we consider is a variation of the standard active “Top- $\kappa$  Selection” problem. Suppose we are presented with a finite set of  $K$  possible alternatives  $\mathcal{A} = \{a_1, a_2, a_3, \dots, a_K\}$  and to each possible pair of alternatives  $(a_i, a_j)$ , there is associated a random variable  $X_{i,j}$  with unknown finite mean  $\mu_{i,j}$ , representing the expected outcome of a “pairwise comparison” of alternative  $a_i$  against alternative  $a_j$ . The quality  $S_i$  of an alternative  $a_i$  is determined by the sum of the means of the R.V.’s corresponding to pairwise comparisons of  $a_i$  against all other alternatives:

$$S_i = \sum_{j \neq i} \mu_{i,j}$$

This is commonly known as the *Borda Score* Borda (1784). We assume that comparing an alternative  $a_i$  to  $a_j$  has the same effect as comparing  $a_j$  to  $a_i$ . Thus, the random variables  $X_{i,j}$  are paired, with  $X_{i,j} = -X_{j,i}$  for value-based sample results and  $X_{i,j} = 1 - X_{j,i}$  for binary preference samples. As such, performing a pairwise comparison of two alternatives will affect the estimates of both of their scores.

This model of defining alternative fitness using the Borda scores of their pairwise comparisons does not explicitly assume the existence of any underlying latent value model for alternatives. However, under reasonable conditions, namely Stochastic Transitivity (see Definition 3.1 below) and Pairwise Distinguishability, the Borda score ranking for alternatives will be identical to the underlying latent ranking, should one exist. For a proof of this property, see Appendix.

The aim is to identify the index set  $\mathcal{I} \subset [K]$  of given size  $\kappa$  containing the highest scoring alternatives, i.e, the solution to the optimization problem:

$$\operatorname{argmax}_{\mathcal{I} \subset [K]: |\mathcal{I}|=\kappa} \sum_{i \in \mathcal{I}} S_i$$

This can be done by iteratively selecting pairs of alternatives  $(a_i, a_j)$  and sampling  $X_{i,j}$ , thereby improving the quality of our estimates of the  $\mu_{i,j}$ ’s that comprise the alternative’s scores. In particular, we are interested in cases where the sampling process is deemed “expensive”, either computationally, or due to the need for real-world interactions, and hence the number of samples we can take is limited. The problem becomes how to iteratively select the next pair to sample to maximise the probability of correctly identifying the optimal subset.

## 3. Related Work

### 3.1. Static Sampling

There is a wide variety of research related to ranking problems based on pairwise information. A number of works (Braverman and Mossel, 2008; Negahban et al., 2012; Hajek et al., 2014) present approaches for generating a

complete ranking on static problem cases. While any method that produces a complete ranking can obviously also be used to identify only the top subset of alternatives, such methods are unlikely to be effective as those specifically designed to do so. For top- $\kappa$  selection specifically, a major approach is the class of spectral ranking methods based on *Rank Centrality*, notably the *Spectral MLE* algorithm proposed in Chen and Suh (2015) and further analysed in Jang et al. (2016). Both consider sets of alternatives with underlying (true) preferences based on the popular *Bradley-Terry-Luce* (BTL) model (Bradley and Terry, 1952). This model assumes the existence of an underlying (unknown) vector of weights,  $(w_1, \dots, w_K)$  that parametrises the true preferences over the set of alternatives  $a_1, \dots, a_K$ . These weights are assumed to determine the probability of each outcome of a pairwise comparison between two alternatives: specifically, in a comparison between alternatives  $a_i$  and  $a_j$ , the probability that  $a_i$  wins is given by  $\frac{w_i}{w_i + w_j}$ .

Recently, Suh et al. (2017) further developed *Rank Centrality* and *Spectral MLE* methods for top- $\kappa$  selection under the BTL model to include “adversarial” settings, where a portion of sample results are deliberately falsified. This is designed to make the methods more robust to real world effects, for example, to the effect of spammers and manipulation of internet survey results.

Although popular, the strong parametric assumptions of the BTL model often fail in real-world applications, (see, for example, Ballinger and Wilcox (1997)). These limitations are discussed in detail in Shah et al. (2016). In that work, the authors suggest the class of *Strongly Stochastically Transitive* (SST) models first defined by Fishburn (1973) to be more consistent with experimental data. This more general class of models is based on the assumption of the SST condition, stated here for binary comparison outcomes:

**Definition 3.1.** *Strong Stochastic Transitivity* condition. Given alternatives  $a_i$ ,  $a_j$  and  $a_k$  with pairwise comparison means  $\mu_{i,j}$ ,  $\mu_{j,k}$  and  $\mu_{i,k}$ , then:

$$\mu_{i,j} \geq \frac{1}{2} \text{ and } \mu_{j,k} \geq \frac{1}{2} \implies \mu_{i,k} \geq \max\{\mu_{i,j}, \mu_{j,k}\}$$

The class of SST models includes the BTL model, as well as other well known parametric models such as the Thurstone model (Thurstone, 1927). Shah and Wainwright (2016) explore methods both for complete ranking and top- $\kappa$  selection for SST models, proposing a simple and computationally efficient counting algorithm based on the Copeland score of each alternative. Chen et al. (2017) also look at static top- $\kappa$  selection with the SST model, presenting a counting algorithm with adaptations to better account for the varying importance of different sample results.

### 3.2. Active Sampling

Where possible, it is often advantageous to actively choose which pairwise comparisons to perform, by sampling sequentially and taking account of previous sample outcomes to choose more relevant or informative pairs to compare. A range of active sampling methods exist that attempt to capture this benefit. For

complete ranking, see Busa-Fekete et al. (2014) for Mallows models, or Maystre and Grossglauser (2015) which discusses *Rank Centrality* and *QuickSort* based approaches for the BTL model. Closely related is the so called *Duelling Bandits* problem described in Yue et al. (2012) and Yue and Joachims (2011) for finding the best single alternative through active pairwise sampling, using an underlying sampling model based on the SST condition. Urvoy et al. (2013) proposes the SAVAGE algorithm, a more general duelling bandit method for identifying the top element without the stochastic transitivity assumption. For a good overview of pairwise learning methods in the context of bandit algorithms, see Busa-Fekete and Hüllermeier (2014).

Several works also explicitly address active pairwise top- $\kappa$  selection. Mohajer and Suh (2016) propose an algorithm to return, in ranked order, the top- $\kappa$  elements of a set. To perform this task, the method relies on partitioning the set of alternatives to be investigated into smaller subsets and constructing binary preference trees on these sets. The method is computationally efficient, requiring only parts of these preference trees to be rebuilt after an alternative is selected for the top set. For the partitioning method to be valid, a certain degree of regularity between the underlying ranking of the items and the individual pairwise comparison means is required, but this assumption is weaker than that of SST. Eriksson (2013) considers a graph-based approach to top- $\kappa$  selection in a setting with uniform noise, which also imposes some consistency conditions, namely the underlying directed preference graphs must be acyclic.

One approach for active top- $\kappa$  selection allowing cyclical preferences is through the use of *Successive Elimination* or *Racing* algorithms, first introduced in Maron and Moore (1994) and Maron and Moore (1997). These iterative methods provide a framework for dealing with sampling uncertainty, designed to replicate a race. During the sampling process, as the quantity of information about each alternative increases, particularly well-performing alternatives can be allowed to “finish early” and are selected, while those that lag behind are eliminated, with further sampling focussed solely on alternatives remaining in the race. Although the standard racing implementation includes the idea of a maximum budget, the proportion of this budget utilized by the race is usually variable, as the algorithm terminates upon reaching a solution set with desired size, having successively eliminated alternatives during the sampling process based on a probabilistic bound based on an accuracy parameter  $\alpha$ . Fixed budget adaptations of the racing framework (see for example Branke and Elomari (2013)) do exist, aiming to adaptively tune the accuracy parameter  $\alpha$  to maximise performance within a given budget constraint.

Racing has been applied to a variety of contexts including model selection and parameter tuning. Heidrich-Meisner and Igel (2009) describe in detail the *selectRace* procedure for obtaining the best  $\mu$  of  $\lambda$  alternatives, using both the Hoeffding and empirical Bernstein bounds. Busa-Fekete et al. (2013b) and Busa-Fekete et al. (2013a) adapt the Hoeffding racing algorithm to the pairwise preference sampling problem, proposing a preference-based racing (PBR) framework with three different sampling strategies. They compare these strategies using real-world football data, and empirically demonstrate their performance

on several synthetic top- $\kappa$  selection problems. Additionally, Heckel et al. (2016) proposes the *Active Ranking (AR)* algorithm, a sequential racing method for dividing the set of alternatives into arbitrary subsets based on Borda Score, and compare the performance of their more general method with other approaches that assume the BTL model or SST, such as the *Duelling Bandit* approach from Yue and Joachims (2011). To our knowledge, Heckel et al. (2016) represents the current state of the art for subset selection using pairwise comparisons with sampling uncertainty and without parametric models like BTL or regularity assumptions such as Stochastic Transitivity.

The *Simulation Optimization* community has developed several alternative approaches to ranking and selection, as discussed in Branke et al. (2007). In particular, the authors describe in detail two classes of Bayesian methods; expected Value of Information Procedures (VIP) and the Optimal Computing Budget Allocation (OCBA).

OCBA refers to a group of procedures first proposed in Chen (1996), and further developed in Chen et al. (2000) and Chen et al. (2010). In Chen et al. (2008), the authors adapt a version of OCBA for optimal subset selection. The problem considered in this work is the classic selection problem, where allocating a simulation run corresponds directly with sampling the score of an alternative. As OCBA is well suited to stochastic simulation optimization problems, and has been shown to be an effective sampling policy for subset selection, it seems reasonable to expect OCBA could be well applied to the pairwise problem.

A popular variant of the VIP approach is the Knowledge Gradient (KG) policy first proposed in Gupta and Miescke (1994) and developed in Frazier et al. (2008) and Chick et al. (2010). The KG policy sequentially samples alternatives based on myopically optimising the expected value of information gained by performing a single additional sample. Frazier et al. (2008) demonstrate that the KG policy is able to perform efficiently where sample measurements are normally distributed. Kamiński (2015) identifies potential limitations of the KG for discrete measurement cases, proposing adapted sample selection methods demonstrated to improve performance in the Bernoulli case. We have reported some preliminary investigation on using OCBA and KG in the context of pairwise comparisons in Groves and Branke (2016). Independent of our work, Priekule and Meisel (2017) show empirically that KG works better than Equal allocation, but can get stuck in a pairwise comparison setting.

## 4. Algorithm Details

In our literature review, we identified two possible sampling policies that can be adapted to address the problem defined in Section 2. In this section, we discuss them in more detail, along with our modifications for pairwise sampling.

### 4.1. Pairwise Optimal Computing Budget Allocation (POCBAm)

Optimal Computing Budget Allocation (OCBA) refers to a class of sampling allocation policies based on a Bayesian framework. Since it was first proposed

in Chen (1996), several different variants of OCBA have been developed (Chen et al., 2000, 2010; Mattila and Virtanen, 2015). The procedure was adapted in Chen et al. (2008) for optimal subset selection, with the name OCBAm to refer to the selection of multiple elements.

Here, we implement the variant first defined by Chen et al. (1997), and later also evaluated by Branke et al. (2007), which we adapt both for selecting a subset rather than a single alternative and to use pairwise comparisons, and thus refer to as POCBAm. At each stage of the sampling process POCBAm aims to maximise the estimated increase in the probability of correct selection (PCS) gained from the sample. To estimate PCS, we consider the information we have gained from our sampling process to far; we have an estimate  $\tilde{\mu}_{i,j}$  for the mean of each sample outcome  $\mu_{i,j}$ , and the standard deviation  $\tilde{\sigma}_{i,j}$  of the sampling results obtained so far. Using a Gaussian approximation, we model the uncertainty of our estimate  $\tilde{\mu}_{i,j}$  using the standard error  $\frac{\tilde{\sigma}_{i,j}}{\sqrt{n_{i,j}}}$ , where  $n_{i,j}$  denotes the number of samples performed of the pair  $(a_i, a_j)$ . We use these pairwise estimates to construct distributions  $\tilde{\mathcal{S}}_p$  for the overall Borda scores of each alternative  $a_p$ :

$$\tilde{\mathcal{S}}_p \sim \mathcal{N} \left( \hat{\boldsymbol{\mu}}_p = \sum_{q, q \neq p} \tilde{\mu}_{p,q}, \hat{\boldsymbol{\sigma}}_p^2 = \sum_{q, q \neq p} \frac{\tilde{\sigma}_{p,q}^2}{n_{p,q}} \right) \quad (4.1.1)$$

With these Borda score distributions, our *expected PCS* (*EPCS*) would simply be the probability that each of the alternative scores does indeed fall in the correct set, i.e:

$$EPCS = \mathbb{P}\{\tilde{\mathcal{S}}_p > \tilde{\mathcal{S}}_q, \text{ for all } p \in \mathcal{I}, q \notin \mathcal{I}\}$$

As we only need the relative values of the *EPCS* for each pair, we use the lower bound *approximate expected probability of correct selection* (*AEPCS*) as described in Chen et al. (2008) to simplify the calculation:

For a constant  $c$ :

$$EPCS \geq \mathbb{P} \left[ \left( \bigcap_{p \in \mathcal{I}} \{\tilde{\mathcal{S}}_p > c\} \right) \cap \left( \bigcap_{q \notin \mathcal{I}} \{\tilde{\mathcal{S}}_q < c\} \right) \right] \equiv AEPCS \quad (4.1.2)$$

To obtain the best approximation of EPCS, we want to choose  $c$  in order to maximise AEPCS and make our lower bound as tight as possible. As suggested in Chen et al. (2008), we use:

$$c = \frac{\hat{\sigma}_{\kappa+1} \hat{\boldsymbol{\mu}}_{\kappa} + \hat{\sigma}_{\kappa} \hat{\boldsymbol{\mu}}_{\kappa+1}}{\hat{\sigma}_{\kappa} + \hat{\sigma}_{\kappa+1}}$$

where  $\hat{\boldsymbol{\mu}}_{\kappa}$ ,  $\hat{\sigma}_{\kappa}$  and  $\hat{\boldsymbol{\mu}}_{\kappa+1}$ ,  $\hat{\sigma}_{\kappa+1}$  are the score means and standard errors of the alternatives currently ranked  $\kappa^{\text{th}}$  and  $(\kappa + 1)^{\text{th}}$  respectively.



However, calculating this probability directly is not straightforward. Unlike Chen et al. (2008), alternative scores are not independent as the sums used to calculate  $\tilde{\mathcal{S}}_p$  and  $\tilde{\mathcal{S}}_q$  include the mean estimates  $\mu_{p,q}$  and  $\mu_{q,p}$  of the paired random variables  $X_{p,q}$  and  $X_{q,p}$  respectively, as described above. Instead, this pairing ensures that the correlation  $\rho_{p,q}$  is negative between any pair of alternatives. Thus using Slepian's Theorem, as described in Tong (1980) (Theorem 2.1.1 and Corollary 1), and assuming joint normality between alternatives, we can produce an upper bound for the parts of *AEPCS* from the sets of alternatives on each side of the threshold:

$$\begin{aligned}\mathbb{P}\left(\bigcap_{p \in \mathcal{I}} \{\tilde{\mathcal{S}}_p > c\}\right) &\leq \prod_{p \in \mathcal{I}} \mathbb{P}\{\tilde{\mathcal{S}}_p > c\} \\ \mathbb{P}\left(\bigcap_{q \notin \mathcal{I}} \{\tilde{\mathcal{S}}_q < c\}\right) &\leq \prod_{q \notin \mathcal{I}} \mathbb{P}\{\tilde{\mathcal{S}}_q < c\}\end{aligned}\tag{4.1.3}$$

Sketch proofs of inequalities 4.1.3 are given in the Appendix. Similarly, the negative correlations between alternative scores gives us an obvious lower bound for pairs of alternatives from either side of the threshold. For  $p \in \mathcal{I}, q \notin \mathcal{I}$ :

$$\mathbb{P}(\tilde{\mathcal{S}}_p > c) \mathbb{P}(\tilde{\mathcal{S}}_q < c) \leq \mathbb{P}(\{\tilde{\mathcal{S}}_p > c\} \cap \{\tilde{\mathcal{S}}_q < c\})\tag{4.1.4}$$

Given these bounds, it seems reasonable to expect that the product over elements of Equation 4.1.2 will provide an acceptable and easy to calculate approximation of *AEPCS*, given again that only relative values are needed. Thus, in practice we approximate *AEPCS* using:

$$AEPCS \approx \prod_{p \in \mathcal{I}} \mathbb{P}\{\tilde{\mathcal{S}}_p > c\} \prod_{q \notin \mathcal{I}} \mathbb{P}\{\tilde{\mathcal{S}}_q < c\}\tag{4.1.5}$$

To estimate the expected increase in *AEPCS* due to allocating an additional sample, the POCBAM procedure considers the effect of allocating a single additional sample to a particular pairwise comparison and none to the others. The expectation is that, by collecting an additional sample from the random variable corresponding to that pair, the estimate of the sample mean and standard deviation will not change (as they are calculated using unbiased estimates), but the standard error of our estimate of the mean of the outcome from that pairwise comparison will decrease. We model this effect, for a sample allocated to the pair  $(a_i, a_j)$ , by scaling the distribution of the scores  $\tilde{\mathcal{S}}_p$  of the alternatives  $a_p$  to our expected post-sample distributions  $\tilde{\mathcal{S}}_p^{i,j}$ :

$$\tilde{\mathcal{S}}_p^{i,j} \sim \mathcal{N}\left(\hat{\mu}_p^{i,j} = \sum_{q, q \neq p} \tilde{\mu}_{p,q}, (\hat{\sigma}_p^{i,j})^2 = \sum_{q, q \neq p} \frac{\tilde{\sigma}_{p,q}^2}{n_{p,q} + \mathbb{I}\{p, q = i, j\}}\right)\tag{4.1.6}$$

where  $\mathbb{I}\{p, q = i, j\}$  is the indicator function that returns 1 if either  $p = i$  and  $q = j$ , or  $q = i$  and  $p = j$ . Calculating  $\tilde{\mathcal{S}}_p^{i,j}$  for all alternatives allows us to obtain

a prediction for *AEPCS* after having performed the additional comparison of  $(a_i, a_j)$ :

$$\text{For } c = \frac{\hat{\sigma}_{\kappa+1}\hat{\mu}_{\kappa} + \hat{\sigma}_{\kappa}\hat{\mu}_{\kappa+1}}{\hat{\sigma}_{\kappa} + \hat{\sigma}_{\kappa+1}},$$

$$\begin{aligned} AEPCS^{i,j} &= \prod_{p \in \mathcal{I}} \mathbb{P}\{\tilde{\mathcal{S}}_p^{i,j} > c\} \prod_{q \notin \mathcal{I}} \mathbb{P}\{\tilde{\mathcal{S}}_q^{i,j} < c\} \\ &= \prod_{p \in \mathcal{I}} \left(1 - \Phi\left(\frac{c - \hat{\mu}_p^{i,j}}{\hat{\sigma}_p^{i,j}}\right)\right) \prod_{q \notin \mathcal{I}} \left(\Phi\left(\frac{c - \hat{\mu}_q^{i,j}}{\hat{\sigma}_q^{i,j}}\right)\right) \end{aligned} \quad (4.1.7)$$

where  $\Phi$  is the cumulative distribution function of the standard normal distribution.

At each step, POGBAm selects and performs a single sample of the pair that maximises  $AEPCS^{i,j}$ , before recalculating, repeating until the pre-sample  $AEPCS > (1 - \alpha)$  for a pre-specified accuracy parameter  $\alpha \in [0, 1]$ . Initial values for pairwise sample mean and variance estimates  $\hat{\mu}_{i,j}$  and  $\hat{\sigma}_{i,j}^2$  are obtained by performing an initial warm-up phase where each alternative pair is sampled  $n_0$  times. With discrete sample values, we apply add-one Laplace smoothing Lidstone (1920) to our initial estimates to ensure  $\hat{\sigma}_{i,j}^2 > 0$ . This is necessary to guarantee asymptotic optimality, discussed further in Section 4.3 below.

#### 4.2. Pairwise Knowledge Gradient (PKG)

Pairwise Knowledge Gradient (PKG) is a one-step Bayesian look-ahead policy that aims to maximise the expected value gained by collecting one additional sample under the assumption that the sampling process will terminate immediately afterwards. In the context of optimal subset selection, given an index set  $\mathcal{I}$ , its value is typically determined by the zero-one loss function:

$$U(\mathcal{I}) = \begin{cases} 1, & \text{if } \mathcal{I} \text{ is correct} \\ 0, & \text{otherwise} \end{cases}$$

Suppose during the sampling process, we currently consider the index set  $\mathcal{I}$  to contain the  $\kappa$  best alternatives and denote the (as yet unknown) best index set we would obtain after sampling a pair  $(a_i, a_j)$  by  $\mathcal{I}^{i,j}$ . The expected value gain of such a sample is simply:

$$V^{i,j} = \mathbb{P}\{U(\mathcal{I}^{i,j}) = 1 | U(\mathcal{I}) = 0\} - \mathbb{P}\{U(\mathcal{I}^{i,j}) = 0 | U(\mathcal{I}) = 1\}$$

However, as we do not know the value of  $U(\mathcal{I})$  during our sampling process,  $V^{i,j}$  cannot be computed. To allow us to approximate it, we make the assumption that the information gained by further sampling should improve our ability to identify the correct index set and thus will not cause us to discard a correct index set  $\mathcal{I}$ , i.e that  $\mathbb{P}\{U(\mathcal{I}^{i,j}) = 0 | U(\mathcal{I}) = 1\} = 0$ . Under this assumption, and the assumption that the next sample will be the last, the expected value of information gained from performing a sample is simply the probability that

---

**POCBAm Procedure**


---

**INPUT:** Set of  $K$  alternatives  $\{a_1, \dots, a_K\}$ ,  
 Required selection size  $\kappa$ ,  
 Accuracy parameter  $\alpha$ .  
**INITIALIZE:** Perform  $n_0$  samples of each pair of alternatives;  
 $n_{p,q} = n_0$  for all  $p, q$ ,  
 Sample means  $\tilde{\mu}_{p,q} = \frac{1}{n_{p,q}} \sum X_{p,q}$ , and:  
 Standard dev.  $\tilde{\sigma}_{p,q} = \sqrt{\frac{1}{n_{p,q}-1} \sum (X_{p,q} - \tilde{\mu}_{p,q})^2}$ ,  
 For all  $p = 1, \dots, K$ : alternative scores  $S_p = \sum_{q,q \neq p} \tilde{\mu}_{p,q}$ ,  
 Index set  $\mathcal{I}$  of best  $\kappa$  alternatives.  
  
**WHILE**  $AEPCS < (1 - \alpha)$  **DO:**  
   **FOR ALL PAIRS**  $(a_i, a_j)$ :  
     **UPDATE:**  
       For all  $\mathbf{p} = 1, \dots, K$ :  
         Alternative score means  $\hat{\boldsymbol{\mu}}_{\mathbf{p}}^{i,j} := S_p$ ,  
         Alternative score std. devs.  $\hat{\boldsymbol{\sigma}}_{\mathbf{p}}^{i,j} := \sum_{q,q \neq p} \frac{\tilde{\sigma}_{p,q}^2}{n_{p,q} + \mathbb{1}\{p,q=i,j\}}$ ,  
         Boundary value  $c = \frac{\tilde{\sigma}_{\kappa+1} \hat{\boldsymbol{\mu}}_{\kappa} + \hat{\boldsymbol{\sigma}}_{\kappa} \hat{\boldsymbol{\mu}}_{\kappa+1}}{\hat{\boldsymbol{\sigma}}_{\kappa} + \hat{\boldsymbol{\sigma}}_{\kappa+1}}$ ,  
          $AEPCS^{i,j} = \prod_{p \in \mathcal{I}} \left( 1 - \Phi \left( \frac{c - \hat{\boldsymbol{\mu}}_p^{i,j}}{\hat{\boldsymbol{\sigma}}_p^{i,j}} \right) \right) \prod_{q \notin \mathcal{I}} \left( \Phi \left( \frac{c - \hat{\boldsymbol{\mu}}_q^{i,j}}{\hat{\boldsymbol{\sigma}}_q^{i,j}} \right) \right)$ .  
     **END FOR**  
     **SAMPLE:** Select pair  $(a_i, a_j)$  that maximises  $AEPCS^{i,j}$ ,  
       Perform sample of  $(a_i, a_j)$ ,  
        $n_{i,j} \leftarrow n_{i,j} + 1$ ,  
       **UPDATE:**  $\tilde{\mu}_{i,j}, \tilde{\sigma}_{i,j}, S_i, S_j$ ,  
     **UPDATE:**  $\mathcal{I}$ .  
**END WHILE**  
  
**RETURN**  $\mathcal{I}$

---

the sample will change our estimate of the index set. Thus, we define the approximate value gain  $AV^{i,j}$  of sampling the pair  $(a_i, a_j)$ :

$$AV^{i,j} := \mathbb{P}\{\mathcal{I}^{i,j} \neq \mathcal{I}\} \quad (4.2.1)$$

For the sample to change our current index set  $\mathcal{I}$ , these score changes must be sufficiently large to move one of  $S_i, S_j$  either into, or out of the current  $\kappa$  best score estimates. The sample may either (i) increase the score estimate  $S_i$  of  $a_i$ , and thus decrease  $S_j$  by an equal amount, or (ii) decrease  $S_i$  and so increase  $S_j$  by a corresponding amount. For case (i), we denote the required increase in  $S_i$  to change  $\mathcal{I}$  by  $\delta_i^{i,j}$ . Similarly, for case (ii), we denote the required increase in  $S_j$  to change  $\mathcal{I}$  by  $\delta_j^{i,j}$ . There are several cases for calculating  $\delta_i^{i,j}$  and  $\delta_j^{i,j}$ , dependent on whether  $a_i$  and  $a_j$  are present in the current estimated index set  $\mathcal{I}$ :

$$\delta_i^{i,j} = \begin{cases} S_\kappa - S_i & \text{if } a_i, a_j \notin \mathcal{I} \\ S_j - S_{\kappa+1} & \text{if } a_i, a_j \in \mathcal{I} \\ \min\{\frac{S_j - S_i}{2}, S_\kappa - S_i, S_j - S_{\kappa+1}\} & \text{if } a_j \in \mathcal{I}, a_i \notin \mathcal{I} \\ \infty & \text{if } a_i \in \mathcal{I}, a_j \notin \mathcal{I} \end{cases} \quad (4.2.2)$$

and vice versa for  $\delta_j^{i,j}$ . If both  $a_i$  and  $a_j$  are outside the current index set, the increase in  $S_i$  must be sufficiently large to make  $S_i$  exceed  $S_\kappa$ , the lowest score for alternatives currently in  $\mathcal{I}$ . Similarly, if  $a_i$  and  $a_j$  are both currently in  $\mathcal{I}$ , the increase in  $S_i$  must be large enough that the corresponding decrease in  $S_j$  is enough to reduce  $S_j$  to below  $S_{\kappa+1}$ , the highest score for alternatives not in  $\mathcal{I}$ . Alternatively, if  $a_i \notin \mathcal{I}, a_j \in \mathcal{I}$ ,  $S_i$  must either increase enough to exceed  $S_\kappa$ , or cause a decrease in  $S_j$  sufficiently large to reduce it below  $S_{\kappa+1}$ , or both increase  $S_i$  and decrease  $S_j$  enough to make  $S_i > S_j$ . Finally, if  $a_i \in \mathcal{I}, a_j \notin \mathcal{I}$ , no increase in  $S_i$  can change  $\mathcal{I}$  so  $\delta_i^{i,j}$  is infinite. However, in this case,  $\delta_j^{i,j} = \min\{\frac{S_i - S_j}{2}, S_\kappa - S_j, S_i - S_{\kappa+1}\}$ , so at least one of  $\delta_i^{i,j}$  and  $\delta_j^{i,j}$  will always be finite. The sampling outcome required to change our estimate of  $\mu_{i,j}$  from  $\tilde{\mu}_{i,j}$  after  $n_{i,j}$  samples, to  $\tilde{\mu}_{i,j} + \delta_i^{i,j}$  after  $n_{i,j} + 1$  samples, and thereby increase  $S_i$  by  $\delta_i^{i,j}$ , is then simply:

$$\Delta_i^{i,j} = \delta_i^{i,j}(n_{i,j} + 1) + \tilde{\mu}_{i,j} \quad (4.2.3)$$

So a sample result from the pair  $(a_i, a_j)$  of at least  $\Delta_i^{i,j}$ , or at least  $\Delta_j^{i,j}$  will cause our index set  $\mathcal{I}$  to change. Thus, the expected value of information  $AV^{i,j}$  from this sample under our knowledge gradient assumptions is just the probability of either of the required sampling outcomes. With Gaussian sampling noise, this is given by:

$$AV^{i,j} = 2 - \left[ \Phi\left(\frac{|\Delta_i^{i,j} - \tilde{\mu}_{i,j}|}{\tilde{\sigma}_{i,j}}\right) + \Phi\left(\frac{|\Delta_j^{i,j} - \tilde{\mu}_{i,j}|}{\tilde{\sigma}_{i,j}}\right) \right] \quad (4.2.4)$$

where  $\Phi$  is the cumulative distribution function of the standard normal distribution. At each step after our initial warm-up phase, we choose to perform the sample that maximises  $AV$ .

Throughout our sampling process, we maintain estimates of sampling mean  $\tilde{\mu}_{i,j}$  and variance  $\tilde{\sigma}_{i,j}^2$  for each pair, which we can use to construct an estimate for the probability of correct selection given our sampling results so far. As with the POCBAm method, we use the *AEPCS* approximation given in Section 4.1 to simplify the calculation, differing only in that we do not scale the variance of our alternative score distributions  $\tilde{S}_p$  to predict future sample effects as we do with POCBAm, instead using only the actual sampling results obtained so far. We use *AEPCS* as a stopping criterion, halting our sampling process when  $AEPCS > (1 - \alpha)$ .

#### 4.3. Asymptotic Optimality

A desirable property for sampling methods is that of *Asymptotic Optimality*; the guarantee of convergence to the best possible solution given an infinite sampling budget. For instance, the simple policy of uniform sample allocation is asymptotically optimal. Under this policy, with an infinite sampling budget, infinitely many samples will be allocated to each possible pair and so each pairwise mean estimate and therefore all Borda Score estimates for alternatives will converge to the true value. This idea is important when discussing asymptotic optimality, as so long as we can guarantee our sampling method will eventually allocate infinitely many samples to each pair, the Central Limit Theorem (CLT) justifies that eventually our Borda Score estimates for alternatives will be sufficiently accurate to guarantee we select the correct subset of high scoring alternatives.

The asymptotic optimality of OCBA methods on standard (i.e non-pairwise) ranking and selection problems is well established, with proofs of the property given for different formulations of the method in Frazier and Powell (2008) and Chen and Lee (2011).

**Proposition 1.** Pairwise OCBA for Top-k selection (POCBAm) is asymptotically optimal.

*Proof.* After our  $n_0$  warm-up samples of each pair we have  $\tilde{\sigma}_{i,j} > 0$  for all pairs  $a_i, a_j$  and thus  $AEPCS^{i,j} > 0$ . Now, when we sample a pair, we only affect the pairwise mean and score estimates of the two alternatives directly involved in the comparison, leaving most of the terms in *AEPCS* unchanged. Thus, we can write the expected increase in *AEPCS* due to sampling  $(a_i, a_j)$  (which we denote  $\Delta AEPCS^{i,j}$ ) by:

$$\Delta AEPCS^{i,j} = C_{i,j} \left[ \Phi \left( \frac{(|c - \tilde{\mu}_{i,j}|)(n_{i,j} + 1)}{\tilde{\sigma}_{i,j}} \right) - \Phi \left( \frac{(|c - \tilde{\mu}_{i,j}|)n_{i,j}}{\tilde{\sigma}_{i,j}} \right) \right]$$

Where  $\Phi$  is the cumulative distribution function of the standard normal distribution. With  $0 < C_{i,j} < 1$  being the product of all terms in  $AEPCS^{i,j}$  that

---

**PKG Procedure**


---

INPUT: Set of  $K$  alternatives  $\{a_1, \dots, a_K\}$ ,  
 Required selection size  $\kappa$ ,  
 Accuracy parameter  $\alpha$ .

INITIALIZE: Perform  $n_0$  samples of each pair of alternatives;  
 $n_{p,q} = n_0$  for all  $p, q$ , and:  
 Sample means  $\tilde{\mu}_{p,q} = \frac{1}{n_{p,q}} \sum X_{p,q}$ ,  
 Standard dev.  $\tilde{\sigma}_{p,q} = \sqrt{\frac{1}{n_{p,q}-1} \sum (X_{p,q} - \tilde{\mu}_{p,q})^2}$ ,  
 For all  $p = 1, \dots, K$ : alternative scores  $S_p = \sum_{q,q \neq p} \tilde{\mu}_{p,q}$ ,  
 Index set  $\mathcal{I}$  of best  $\kappa$  alternatives.

WHILE  $AEPCS < (1 - \alpha)$  DO:

FOR ALL PAIRS  $(a_i, a_j)$ :

UPDATE: Required score changes  $\delta_i^{i,j}$  and  $\delta_j^{i,j}$ ,  
 $\Delta_i^{i,j} = \delta_i^{i,j}(n_{i,j} + 1) + \tilde{\mu}_{i,j}$ ,  
 $\Delta_j^{i,j} = \delta_j^{i,j}(n_{j,i} + 1) + \tilde{\mu}_{j,i}$ ,  
 $AV^{i,j} = 2 - \left[ \Phi\left(\frac{|\Delta_i^{i,j} - \tilde{\mu}_{i,j}|}{\tilde{\sigma}_{i,j}}\right) + \Phi\left(\frac{|\Delta_j^{i,j} - \tilde{\mu}_{j,i}|}{\tilde{\sigma}_{j,i}}\right) \right]$ .

END FOR

SAMPLE: Select pair  $(a_i, a_j)$  that maximises  $AV^{i,j}$ ,  
 If  $\max_{i,j} \{AV^{i,j}\} = 0$ , select sample uniformly at random,  
 Perform sample of  $(a_i, a_j)$ ,  
 $n_{i,j} \leftarrow n_{i,j} + 1$ ,  
 UPDATE:  $\tilde{\mu}_{i,j}, \tilde{\sigma}_{i,j}, S_i, S_j$ ,

EST. PCS:

For all  $\mathbf{p} = 1, \dots, K$ :

Alternative score means  $\hat{\mu}_{\mathbf{p}} := S_{\mathbf{p}}$ ,

Alternative score std. devs.  $\hat{\sigma}_{\mathbf{p}} := \sum_{q,q \neq \mathbf{p}} \tilde{\sigma}_{\mathbf{p},q}^2$

Boundary value  $c = \frac{\hat{\sigma}_{\kappa+1} \hat{\mu}_{\kappa} + \hat{\sigma}_{\kappa} \hat{\mu}_{\kappa+1}}{\hat{\sigma}_{\kappa} + \hat{\sigma}_{\kappa+1}}$ ,

$AEPCS = \prod_{p \in \mathcal{I}} \left(1 - \Phi\left(\frac{\hat{\mu}_{\mathbf{p}} - c}{\hat{\sigma}_{\mathbf{p}}}\right)\right) \prod_{q \notin \mathcal{I}} \left(\Phi\left(\frac{c - \hat{\mu}_{\mathbf{q}}}{\hat{\sigma}_{\mathbf{q}}}\right)\right)$ .

UPDATE:  $\mathcal{I}$ .

END WHILE

RETURN  $\mathcal{I}$

---

are unaffected by sampling  $(a_i, a_j)$ . Now:

$$\lim_{n_{i,j} \rightarrow \infty} \Delta AEPCS^{i,j} = C_{i,j}(\Phi(+\infty) - \Phi(+\infty)) = C_{i,j}(1 - 1) = 0$$

Therefore, as our total number of samples allocated  $N = \sum_{\forall i,j} n_{i,j} \rightarrow \infty$ , for at least some pairs  $(a_i, a_j)$ , we must have  $n_{i,j} \rightarrow \infty$  and therefore  $\Delta AEPCS^{i,j} \rightarrow 0$ . If this is the case for all pairs, then we are done. Let  $F$  denote the set of pairs for which  $n_{i,j}$  remains finite, then eventually we must reach a state where we allocate no further samples to  $F$ . But if  $n_{i,j} \rightarrow \infty$  for all pairs not in  $F$ , then for any  $\epsilon > 0$  there exists some number of samples  $N'$  such that, once at least  $N'$  samples have been taken we have  $\max_{(a_i, a_j) \notin F} \Delta AEPCS^{i,j} < \epsilon$ . If we choose  $\epsilon < \min_{(a_i, a_j) \in F} \Delta AEPCS^{i,j}$  then for some  $N^*$  we have  $\max_{(a_i, a_j) \notin F} \Delta AEPCS^{i,j} < \min_{(a_i, a_j) \in F} \Delta AEPCS^{i,j}$  after  $N^*$  samples, and so POCBAM will allocate our next sample to  $F$ . Thus, by contradiction,  $F$  is empty.  $\square$

The question of asymptotic optimality for pairwise knowledge gradient is less straightforward. With unbounded sample responses of finite variance, KG is asymptotically optimal (Frazier and Powell (2008)) and we show in Proposition 2 that this remains true for PKG. However, with bounded sample outcomes, this does not hold, as we discuss below. In Proposition 3, we give a specific example of how the asymptotic optimality of PKG may break with binary sampling.

**Proposition 2.** Pairwise Knowledge Gradient (PKG) is asymptotically optimal for sampling models with unbounded sample results of finite variance.

*Proof.* For every pair  $i, j$  at least one of  $\delta_i^{i,j}$  and  $\delta_j^{i,j}$  must be finite. Thus, at least one of  $\Delta_i^{i,j}$  and  $\Delta_j^{i,j}$  will be finite and, as sample results are unbounded, either  $\mathbb{P}[X_{i,j} > \Delta_i^{i,j}] > 0$  or  $\mathbb{P}[X_{j,i} > \Delta_j^{i,j}] > 0$ . Hence,

$$AV^{i,j} > 0$$

for every pair. Now, as  $\lim_{n_{i,j} \rightarrow \infty} \Delta_i^{i,j} = \infty$ ,

$$\lim_{n_{i,j} \rightarrow \infty} AV^{i,j} = 2 - \left[ \Phi\left(\frac{|\infty - \mu_{i,j}|}{\sigma_{i,j}}\right) + \Phi\left(\frac{|\infty - \mu_{i,j}|}{\sigma_{i,j}}\right) \right] = 2 - (1 + 1) = 0$$

Thus, suppose that as  $N = \sum_{\forall i,j} n_{i,j} \rightarrow \infty$  there are some pairs sampled only finitely many times, and denote these by  $F$ , but again, if  $n_{i,j} \rightarrow \infty$  for all pairs not in  $F$ , then for any  $\epsilon > 0$  there is a number of samples  $N'$  after which  $\max_{(a_i, a_j) \notin F} AV^{i,j} < \epsilon$ . If we choose  $\epsilon < \min_{(a_i, a_j) \in F} AV^{i,j}$  then after some  $N^*$  samples, we have  $\max_{(a_i, a_j) \notin F} AV^{i,j} < \min_{(a_i, a_j) \in F} AV^{i,j}$  and thus PKG will allocate our next sample to  $F$ .  $\square$

However, when sample outcomes are bounded, even the standard formulation of KG can encounter states where it is unable to allocate a sample. At each step of this algorithm and for each pair  $(a_i, a_j)$ , we calculate an estimate  $AV^{i,j}$  of

the probability that collecting a single additional sample will change the scores of alternatives  $a_i$  and  $a_j$  sufficiently to alter our top-rated subset. However, restricting the range from which sample results are drawn limits the change that sampling can make to the alternatives' scores. Specifically, using the example of binary sample outcomes; suppose we are in some knowledge state  $\theta$ , with an estimate of  $\tilde{\mu}_{i,j}^\theta$  for  $X_{i,j}$  (and thus  $\tilde{\mu}_{j,i}^\theta = 1 - \tilde{\mu}_{i,j}^\theta$  for  $X_{j,i}$ ) and we perform a single additional sample of  $X_{i,j}$ . Then we will have:

$$\tilde{\mu}_{i,j}^{\theta+1} = \begin{cases} \tilde{\mu}_{i,j}^\theta + \frac{1}{n_{i,j}^\theta+1}(1 - \tilde{\mu}_{i,j}^\theta), & \text{if } X_{i,j}^{\theta+1} = 1 \\ \tilde{\mu}_{i,j}^\theta - \frac{\tilde{\mu}_{i,j}^\theta}{n_{i,j}^\theta+1}, & \text{if } X_{i,j}^{\theta+1} = 0 \end{cases}$$

and vice versa for  $\tilde{\mu}_{j,i}^\theta$ . If the required difference in estimated score exceeds all these possible change amounts, then no single pairwise sample will be able to alter the current top set. This means that our knowledge gradient values will be:

$$AV^{i,j} = 0, \forall (a_i, a_j)$$

and our Knowledge Gradient policy will be unable to select a sample. This potential problem with the KG policy was hinted at in Powell and Ryzhov (2012), and discussed in detail in Kamiński (2015).

It becomes necessary to consider multiple samples in order to find sampling sequences with non-zero change probabilities. Frazier and Powell (2010) propose the adapted KG(\*) policy, which considers sequences of samples, selecting to perform the sample at the start of the shortest sequence required to change the ranking. They show that this policy performs well, but can be computationally very intensive, as the state space of sampling sequences grows rapidly as the sequences lengthen. In the case of pairwise sampling, with  $\frac{1}{2}(K^2 - K)$  possible sample pairs at each stage, this method rapidly becomes computationally intractable for even modest values of  $K$ .

To solve this, Kamiński (2015) suggests an alternative method for formulating  $AV^{i,j}$ , leading to an adapted policy KG(min), that allocates based on minimising the number of consecutive repeated samples of a single alternative needed to change the selection. For the standard subset selection problem, where simulation directly estimates the score of alternatives, this is sufficient to prevent the policy from failing and restore asymptotic optimality, as for any possible alternative score value  $S$  and accuracy  $\epsilon$ , there is a finite string of sampling outcomes that can move the score estimate of an alternative to within  $\epsilon$  of  $S$ , with non-zero probability. However, this is not necessarily true in the pairwise problem, which we show here for binary sample outcomes. In this example, pairwise outcomes are modelled with Bernoulli random variables  $X_{i,j}$ , paired such that  $X_{i,j} = 1 - X_{j,i}$ . Let  $r_{i,j}$  denote the minimum number of consecutive samples of the pair  $(a_i, a_j)$  required to change the selected subset.

**Proposition 3.** For any  $K > 4$  and with binary pairwise sampling, it is possible that  $r_{i,j} = \infty$  for all  $(a_i, a_j)$  regardless of the selected subset size  $\kappa$



*Proof.* To show this, we aim to construct an example sampling situation whereby  $r_{i,j} < \infty \implies K \leq 4$ . Suppose at some point in our sampling process we have:

$$\tilde{\mu}_{i,j} = \begin{cases} 0.5 & \text{if } i, j \in \mathcal{I} \\ 0.5 & \text{if } i, j \notin \mathcal{I} \\ 1 & \text{if } i \in \mathcal{I}, j \notin \mathcal{I} \\ 0 & \text{if } i \notin \mathcal{I}, j \in \mathcal{I} \end{cases}$$

then the estimated difference in score of the  $\kappa^{th}$  and  $(\kappa + 1)^{th}$  best alternatives will be:

$$\begin{aligned} S_\kappa - S_{\kappa+1} &= \sum_{j, j \neq \kappa} \tilde{\mu}_{\kappa,j} - \sum_{j, j \neq \kappa+1} \tilde{\mu}_{\kappa+1,j} \\ &= (0.5(\kappa - 1) + K - \kappa) - 0.5(K - \kappa - 1) \\ &= 0.5K \end{aligned}$$

Now,  $\min_{i,j} \{\delta_i^{i,j}\} = \delta_{\kappa+1}^{\kappa+1,\kappa} = \frac{1}{2}(S_\kappa - S_{\kappa+1}) = \frac{K}{4}$ , so finitely many samples must be able to alter  $\tilde{\mu}_{\kappa+1,\kappa}$  by at least  $\frac{K}{4}$  for  $r_{\kappa,\kappa+1}$  to be finite. As  $\tilde{\mu}_{i,j} \in [0, 1]$  for all  $a_i, a_j$ , the maximum change in  $\tilde{\mu}_{\kappa+1,\kappa}$  we can obtain is 1. Hence, we require:

$$1 \geq \frac{K}{4} \quad \square$$

Proposition 3 means that, when our sampling process returns binary preference feedback, and we have more than 4 alternatives to choose from, it can be the case that infinitely many samples of a particular pair cannot change the ranking. This means that a pairwise adaptation of KG(min) procedure proposed in Kamiński (2015) can fail, even in the asymptotic limit, which explains the observations reported in Groves and Branke (2016) and Priekule and Meisel (2017). To prevent this asymptotic failure for our PKG policy, we adapt PKG to allow random sample selection in the case that  $AV^{i,j} = 0$  for all pairs  $(a_i, a_j)$ .

## 5. Empirical Testing

In this section, we empirically evaluate the performance of our algorithms against other sampling methods. For comparison, we choose the *Active Ranking (AR)* method from Heckel et al. (2016) and the *Hoeffding Racing (H-Race)* method from Busa-Fekete et al. (2013a). To the best of our knowledge, these methods represent the current state of the art for top- $\kappa$  selection for models without systematic regularity assumptions such as SST. We also include the performance of uniform sample allocation as an additional benchmark. We test their performance on a range of standard scoring models (BTL, SST, Unstructured), with both binary preference and value-based sample results.

### 5.1. Top- $\kappa$ Selection (2 from 5)

Here we simulate the problem of selecting the top 2 alternatives from a set of 5. Pairwise outcomes are binary, i.e for alternatives  $(a_i, a_j)$ ,  $X_{i,j}$  is Bernoulli distributed. We consider three different scoring models:

- BTL model: Here the underlying “true” quality of our alternatives is parametrised by a score vector  $T = (t_{a_1}, \dots, t_{a_5})$ .  $T$  fully determines the matrix of pairwise comparison outcome probabilities with  $\mu_{i,j} = \frac{t_i}{t_i + t_j}$ . We set  $T = (0.9, 0.7, 0.5, 0.3, 0.1)$ .
- SST model: We generate the pairwise comparison probability matrix according to the “Independent Bands” SST model from Shah et al. (2016). As they describe, the class of SST Bernoulli scoring models is characterised up to permutation of elements by the set of matrices whose upper-triangular entries lie in  $[0.5, 1.0]$ , increase along rows and decrease down columns. Thus, we generate the matrix of true comparison means  $M$  first by selecting the entry  $M_{0,1} = \mu_{0,1}$  uniformly at random from  $[\frac{1}{2}, 1]$ , before populating the remainder of the upper triangle of the matrix row-wise, at each stage selecting values uniformly from the allowable interval, i.e bounded above either by 1 or the entry above, and bounded below either by  $\frac{1}{2}$  or the entry to the left.
- Unstructured model: In this model pairwise comparison means are uncorrelated. Each entry in the upper triangle of the comparison matrix  $M$  is sampled independently and uniformly at random from  $[0, 1]$ . The “true” ranking of the alternatives is then determined by their Borda score.

The POCBAm, PKG and AR methods contain an accuracy parameter  $\alpha$  related to stopping time, which we vary to obtain a range of values. For the H-Race method, there are two parameters that affect the width of the confidence interval used to eliminate alternatives from the race, and therefore stopping time:  $\alpha$  and  $n_{max}$ , the maximum number of samples allowed of each particular pair. Specifically, the interval is defined for each alternative pair as:

$$\tilde{\mu}_{i,j} \pm \sqrt{\frac{1}{2n_{i,j}} \log \left( \frac{2K^2 n_{max}}{\alpha} \right)}$$

Thus for the H-Race method we obtain a range of different stopping times by varying  $n_{max}$  for three different values of  $\alpha$ : 1.0, 0.1 and 0.01, and display the best performance from the three. The parameter values used for each method are given in the figure caption for each scoring model. We also included a fixed maximum budget total constraint of 10,000 samples per run for each method to ensure timely completion. Performance for each method is measured by the proportion of top- $\kappa$  subsets correctly identified (success rate) over 10,000 different replications (random seeds).

Figure 1a shows the performance of each method at selecting the top 2 of 5 alternatives for the BTL model scenario. POCBAm is the best performer, with

both PKG and POCBAM outperforming the comparison methods, achieving the same success rate using fewer samples. Both racing methods, particularly the H-Race, struggled due to the loose width of the bounds used to construct their confidence intervals. Without being able to successfully eliminate alternatives from the race before reaching  $n_{max}$  samples of each pair, H-Race performs essentially the same samples as simple uniform allocation.

Figure 1b shows the results for the SST scoring model. Overall the SST scoring model produces easier top- $\kappa$  selection problems than the BTL model used in the first scenario. The method for generating the underlying comparison matrix for the SST scoring model will, on average, produce pairwise means further from 0.5 than for the BTL model. For example, the expectation of the mean  $\mu_{2,3}$  of the alternatives on the boundary between the top subset and the discarded subset will be:

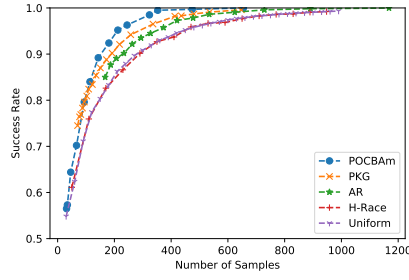
$$\mathbb{E}[\mu_{2,3}] = \frac{0.5 + \mathbb{E}[\mu_{0,2}]}{2} = \frac{0.5 + \frac{\mathbb{E}[\mu_{0,1}] + 1}{2}}{2} = \frac{0.5 + \frac{0.5 + 1}{2} + 1}{2} = 0.6875,$$

compared to  $\frac{0.7}{0.7+0.5} = 0.583$  in the BTL model experiment. Consequently, we see higher success rates at each given budget when compared to Figure 1a, and a much clearer improvement over uniform sampling for the AR method. The larger the differences in Borda score between alternatives, the easier it is for methods to identify which sampling pairs are irrelevant and thus to sample more efficiently than uniformly. However, we see that the confidence bound used by the H-Race method is again too loose to reliably eliminate alternatives before reaching  $n_{max}$  samples of each pair, and thus does not make any improvement over uniform allocation.

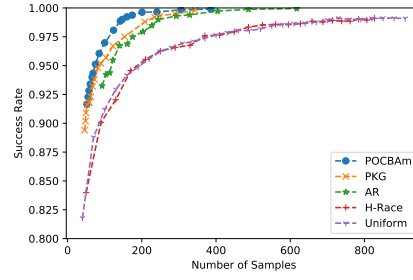
In contrast to the SST scenario, the unstructured scoring model shown in Figure 1c is much harder, with very little average distance between alternative's Borda scores. Both POCBAM and PKG perform well on this problem, with POCBAM achieving the highest success rate as sampling budget increases. To accurately estimate the alternative scores with unstructured preferences, we have to learn far more about the underlying matrix  $M$ . This is particularly difficult for the AR method as, although this method chooses one alternative for the sampling pair directly, the other is selected at random, meaning we would require far more samples to be performed before being sure we have learnt about every entry in  $M$ . This is reflected in low initial performance of AR. The H-Race also struggled in this problem, due to the closeness of the alternative's total scores and the loose confidence bound used, and again fails to improve over the uniform benchmark.

## 5.2. Other System Sizes

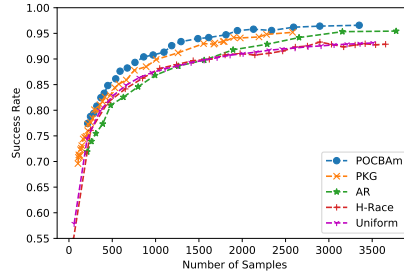
Here we examine effect of the number of alternatives and of the top subset size on the algorithm performance on the SST scoring model used in the previous section. Specifically, we test top 1 of 5 selection (finding the best single element) and top 4 of 10, with the results shown in Figures 2a and 2b respectively. POCBAM is again the best performing method in both cases, reaching



(a) BTL



(b) SST



(c) Unstructured

Figure 1: Performance of POCBA, PKG, AR and H-Race algorithms against random allocation at best 2 of 5 selection for the BTL, SST and unstructured models. For POCBA, we vary  $\alpha$  between 0.5 and 0.01, for PKG between 0.3 and 0.001 and for AR between 0.15 and 0.01. For the H-Race,  $n_{max}$  ranges between 5 and 100 for the BTL and SST models, and between 5 and 450 for the Unstructured model, with  $\alpha = 0.01$  for all three.

perfect accuracy on both problems with substantially fewer samples, particularly on the larger problem. When only needing to identify the single best alternative, both the AR and H-Race methods are able to exclude poorly performing alternatives much earlier as they need only be confident that they are beaten by a single competitor. As sampling budget increases, they are therefore able to allocate the last portion of their sampling budgets more effectively between fewer remaining pairs. This is reflected in their performance, with AR matching PKG and H-Race improving over uniform allocation. Figure 2c shows the performance on a much larger subset selection problem, choosing the top 40 of 100 alternatives. For this problem two other changes were made. Firstly the maximum budget constraint for the variable stopping time methods was increased to 200,000 samples to compensate for the increased number of alternative pairs. Secondly, the PKG method was adapted to use a fixed sampling budget, rather than a variable stopping point based on EPCS. As the number of alternatives increases, but the range of values for each pairwise comparison mean  $\mu_{i,j}$  remains fixed and bounded, the relative effect that changing each sample mean may have on an alternative’s score decreases. This makes it more likely that  $AV^{i,j}$  will fall to zero for some or possibly all alternative pairs, as discussed in Section 4.3. If there are only a few pairs with non-zero  $AV^{i,j}$  values, PKG will only select samples from amongst these, which can prevent EPCS from reaching  $(1 - \alpha)$  even asymptotically. The fixed sampling budget allows PKG to terminate in these cases. If  $AV^{i,j} = 0$  for all pairs, PKG has to resort to random sample allocation. Here we see that these changes limit the effectiveness of PKG, reducing its improvement over uniform allocation compared to the smaller selection problems. Overall, we see that POCBAm is still the best performer. The AR method also performs well, substantially improving over uniform sampling.

### 5.3. Value-based Scoring Models

The final part of this section examines empirical performance of the sampling methods on top 2 of 5 selection on models where pairwise comparison feedback is continuous valued and unbounded. This value-based feedback increases the amount of information received from sampling; instead of simply receiving a 0 – 1 win/loss result as in our previous testing, we now gain a measure of the magnitude of an alternative’s win or loss.

We test performance on value-based SST and unstructured problem models, the BTL model used in the previous section being suitable only for binary preference-based sampling. For both models, we assume that sample outcomes are normally distributed, and choose the underlying variance for each pair uniformly at random from  $[0, 1]$ . The matrix of pairwise comparison means for each model is then generated as described below:

- Value-based SST model: As before, the upper triangle of  $M$  should increase along rows and decrease down columns. Thus we populate the upper triangle of  $M$  using the same procedure as for the binary SST model, except using  $[0, 1]$  instead of  $[0.5, 1]$  as the allowable interval. The lower

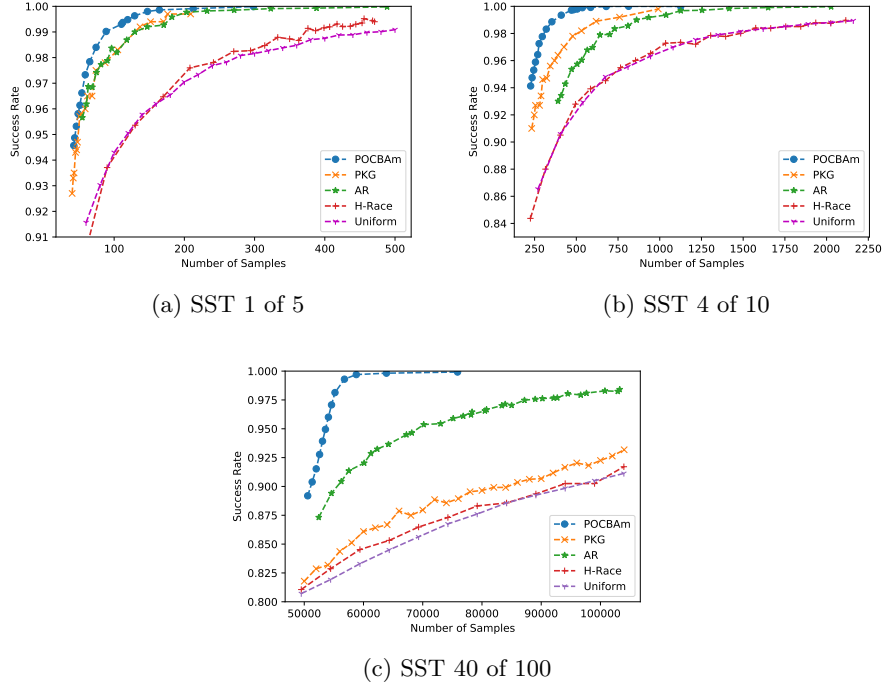


Figure 2: Performance of POCBAm, PKG, AR and H-Race algorithms against random allocation at best 1 of 5 selection (a), best 4 of 10 (b), and best 40 of 100 selection (c), for the SST scoring model. For POCBAm, we vary  $\alpha$  between 0.5 and 0.01, for PKG between 0.3 and 0.001 and for AR between 0.2 and 0.01. For the H-Race, (a) uses  $\alpha = 1.0$ ,  $n_{max}$  between 5 and 100, (b) uses  $\alpha = 0.01$ ,  $n_{max}$  between 5 and 50, and (c) uses  $\alpha = 0.01$ ,  $n_{max}$  between 10 and 25.

Table 1: Percentage reduction in sampling budget to match performance of uniform sample allocation for each sampling method. Best values shown in bold.

		Uniform		Sampling Budget Reduction			
Scoring Model		Samples	Succ. Rate	POCBAm	PKG	AR	H-Race
2 of 5	BTL	1000	0.993	<b>64.7%</b>	35.2%	27.1%	3.34%
	SST	800	0.991	<b>78.0%</b>	70.7%	63.4%	0.512%
	Unstr.	3500	0.931	<b>63.0%</b>	51.9%	24.1%	0.151%
	1 of 5 (SST)	500	0.991	<b>81.7%</b>	71.1%	70.6%	24.8%
	4 of 10 (SST)	2000	0.988	<b>81.0%</b>	66.6%	56.9%	-1.25%
	40 of 100 (SST)	100,000	0.906	<b>49.4%</b>	13.4%	42.5%	-1.48%
Value-Based	SST	200	0.980	63.2%	<b>67.4%</b>	31.4%	25.9%
	Unstr.	2000	0.970	<b>62.4%</b>	60.6%	32.7%	N\A

triangle is then filled according to  $\mu_{j,i} = -\mu_{i,j}$ . Note that value-based SST comparison matrices are skew-symmetric.

- Value-based Unstructured model. Entries in the upper triangle are chosen independently and uniformly at random from  $[0, 1]$ , and the lower triangle filled according to  $\mu_{j,i} = -\mu_{i,j}$ .

Figure 3a shows the performance for the value-based SST model. With normally distributed sample results, PKG will be asymptotically optimal (Proposition 2), so should no longer encounter states where  $AV^{i,j} = 0$  and should no longer have to resort to random samples allocation. As such, performance of PKG and POCBAm seems to be very similar and substantially better than the comparison methods.

We again see that the unstructured model in Figure 3b is much more difficult, with all methods requiring far more samples to reach their stopping points. As with the binary unstructured model from Figure 1c, we see that the AR method initially performs poorly, as its random sample selection is unable to ensure sufficient information collection from all pairs without taking a large number of samples. Interestingly, as the number of samples taken increases, the success rate of the H-Race method falls behind uniform sampling. The bounds for the alternative’s Borda score confidence intervals used by the H-Race are the arithmetic means of the bounds for the individual pairwise confidence bounds, calculated using only the alternatives still included in the race. When pairwise means are uncorrelated, as in the unstructured model, these become progressively poorer estimates of alternatives Borda scores whenever alternatives are removed, leading to incorrect classification of the remaining alternatives.

Table 1 provides a summary of our empirical results, showing the percentage reduction in samples required to achieve the same success rate as uniform sampling for each method on each scoring model.

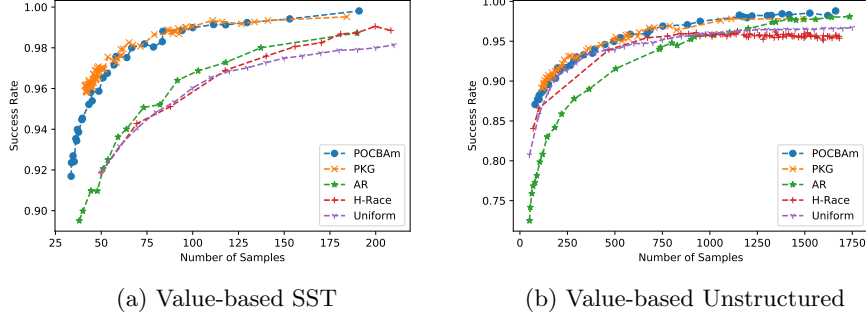


Figure 3: Performance of POCBAm, PKG, AR and H-Race algorithms against random allocation at best 2 of 5 selection with normally distributed sample results for SST (a) and unstructured (b) models. Sub-figure (a) uses  $\alpha$  between 0.5 and 0.01 for POCBAm, between 0.05 and  $10^{-5}$  for PKG and between 0.15 and 0.01 for AR with  $n_{max}$  between 5 and 80, and  $\alpha = 1.0$ . Sub-figure (b) uses  $\alpha$  between 0.3 and 0.001 for POCBAm, between 0.03 and  $10^{-5}$  for PKG, and between 0.15 and 0.001 for AR with  $th\ n_{max}$  between 5 and 2500, and  $\alpha = 0.01$ .

## 6. Conclusion and Future Work

In this work, we have presented two new pairwise subset selection methods, adapted from well known sampling algorithms from the Simulation Optimization community, as well as theoretical guarantees of asymptotically optimal performance under certain conditions. Additionally, we identify an interesting idiosyncrasy of Knowledge Gradient policies with bounded pairwise sampling, where even n-step sampling methods can fail.

In our empirical testing, we see that both PKG and POCBAm offer improvements over current state-of-the-art top- $\kappa$  sampling procedures for scoring models without dependence on structural assumptions like the SST property. POCBAm in particular performed well across all the test scenarios, both with binary and unbounded value-based sample feedback, and with both structured and unstructured underlying models.

A possible future development would be to consider correlations between pairwise sample distributions, as they would occur if there was an underlying (unknown) quality of each solution that would influence the outcome. In such cases, it may be possible to further improve our sampling method by correctly learning and accounting for this dependence between alternative scores.

There are other forms of the standard OCBAm method such as the one described in Chen et al. (2008). These are based on evaluating the asymptotically optimal proportional sample allocation based on current information and then recommending sampling proportionally. As POCBAm was generally the best performing method, it might be interesting to investigate pairwise adaptations of these other OCBA algorithms.



## 7. Acknowledgements

The first author would like to acknowledge funding from the Engineering and Physical Sciences Research Council (EPSRC) through the Mathematics for Real-World Systems CDT at the University of Warwick.

## References

- Ballinger, T., Wilcox, N., 1997. Decisions, error and heterogeneity. *The Economic Journal* 107 (443), 1090–1105.
- Borda, J. C., 1784. Mémoire sur les élections au scrutin. *Histoire de l’Academie Royale des Sciences pour 1781*.
- Bradley, R., Terry, M., 1952. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika* 39 (3/4), 324–345.
- Branke, J., Chick, S., Schmidt, C., 2007. Selecting a selection procedure. *Management Science* 53 (12), 1916–1932.
- Branke, J., Elomari, J., 2013. Racing with a fixed budget and a self-adaptive significance level. In: *Learning and Intelligent Optimization*. Springer, pp. 272–280.
- Braverman, M., Mossel, E., 2008. Noisy sorting without resampling. In: *19th Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, pp. 268–276.
- Busa-Fekete, R., Fober, T., Hüllermeier, E., 2013a. Preference-based evolutionary optimization using generalized racing algorithms. In: *23rd Workshop on Computational Intelligence*; Dortmund. KIT Scientific Publishing, pp. 237–245.
- Busa-Fekete, R., Hüllermeier, E., 2014. A survey of preference-based online learning with bandit algorithms. In: *International Conference on Algorithmic Learning Theory*. Springer, pp. 18–39.
- Busa-Fekete, R., Hüllermeier, E., Szörényi, B., 2014. Preference-based rank elicitation using statistical models: The case of mallows. In: *31st International Conference on Machine Learning*. pp. 1071–1079.
- Busa-Fekete, R., Szorenyi, B., Cheng, W., Weng, P., Hüllermeier, E., 2013b. Top-k selection based on adaptive sampling of noisy preferences. In: *30th International Conference on Machine Learning*. JMLR, pp. 1094–1102.
- Chen, C. H., 1996. A lower bound for the correct subset-selection probability and its application to discrete event simulations. *IEEE Transactions on Automatic Control* 41 (8), 1227–1231.
- Chen, C. H., He, D., Fu, M., Lee, L. H., 2008. Efficient simulation budget allocation for selecting an optimal subset. *Inform Journal on Computing* 20 (4), 579–595.
- Chen, C. H., Lee, L. H., 2011. *Stochastic simulation optimization: an optimal computing budget allocation*. Vol. 1. World scientific.
- Chen, C. H., Lin, J., Yücesan, E., Chick, S., 2000. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dynamic Systems: Theory and Applications* 10 (3), 251–270.

- Chen, C. H., Yücesan, E., Dai, L., Chen, H. C., 2010. Efficient computation of optimal budget allocation for discrete event simulation experiment. *IIE Transactions* 42 (1), 60–70.
- Chen, H. C., Dai, L., Chen, C. H., Yücesan, E., 1997. New development of optimal computing budget allocation for discrete event simulation. In: *Proceedings of the 29th conference on Winter simulation*. IEEE Computer Society, pp. 334–341.
- Chen, X., Gopi, S., Mao, J., Schneider, J., 2017. Competitive analysis of the top-k ranking problem. In: *28th Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, pp. 1245–1264.
- Chen, Y., Suh, C., 2015. Spectral mle: Top-k rank aggregation from pairwise comparisons. In: *32nd International Conference on Machine Learning*. pp. 371–380.
- Chick, S., Branke, J., Schmidt, C., 2010. Sequential sampling to myopically maximize the expected value of information. *Inform Journal on Computing* 22 (1), 71–80.
- Eriksson, B., 2013. Learning to top-k search using pairwise comparisons. In: *16th International Conference on Artificial Intelligence and Statistics*. pp. 265–273.
- Fishburn, P., 1973. Binary choice probabilities: on the varieties of stochastic transitivity. *Journal of Mathematical psychology* 10 (4), 327–352.
- Frazier, P., Powell, W., 2008. Asymptotic optimality of sequential sampling policies for bayesian information collection. Submitted for publication.
- Frazier, P. I., Powell, W. B., 2010. Paradoxes in learning and the marginal value of information. *Decision Analysis* 7 (4), 378–403.
- Frazier, P. I., Powell, W. B., Dayanik, S., 2008. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization* 47 (5), 2410–2439.
- Groves, M., Branke, J., 2016. Optimal subset selection with pairwise comparisons. In: *Proceedings of the DA2PL’2016 EURO Mini Conference*. EURO, pp. 15–20.
- Gupta, S. S., Miescke, K. J., 1994. Bayesian look ahead one stage sampling allocations for selecting the largest normal mean. *Statistical Papers* 35 (1), 169–177.
- Hajek, B., Oh, S., Xu, J., 2014. Minimax-optimal inference from partial rankings. In: *Advances in Neural Information Processing Systems* 27. Curran Associates, Inc., pp. 1475–1483.
- Heckel, R., Shah, N., Ramchandran, K., Wainwright, M., 2016. Active ranking from pairwise comparisons and when parametric assumptions don’t help. *arXiv preprint arXiv:1606.08842*.
- Heidrich-Meisner, V., Igel, C., 2009. Hoeffding and bernstein races for selecting policies in evolutionary direct policy search. In: *26th International Conference on Machine Learning*. ACM, pp. 401–408.
- Jang, M., Kim, S., Suh, C., Oh, S., 2016. Top- $k$  ranking from pairwise comparisons: When spectral ranking is optimal. *arXiv preprint arXiv:1603.04153*.

- Kamiński, B., 2015. Refined knowledge-gradient policy for learning probabilities. *Operations Research Letters* 43 (2), 143–147.
- Lidstone, G. J., 1920. Note on the general case of the bayes-laplace formula for inductive or a posteriori probabilities. *Transactions of the Faculty of Actuaries* 8 (182-192), 13.
- Maron, O., Moore, A., 1994. Hoeffding races: accelerating model selection search for classification and function approximation. *Advances in Neural Information Processing Systems*, 59–66.
- Maron, O., Moore, A., 1997. The racing algorithm: Model selection for lazy learners. *Artificial Intelligence Review* 5 (1), 193–225.
- Mattila, V., Virtanen, K., 2015. Ranking and selection for multiple performance measures using incomplete preference information. *European Journal of Operational Research* 242 (2), 568–579.
- Maystre, L., Grossglauser, M., 2015. Robust active ranking from sparse noisy comparisons. *arXiv preprint arXiv:1502.05556*.
- Mohajer, S., Suh, C., 2016. Active top-k ranking from noisy comparisons. In: *54th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, pp. 875–882.
- Negahban, S., Oh, S., Shah, D., 2012. Iterative ranking from pair-wise comparisons. In: *Advances in Neural Information Processing Systems* 25. Curran Associates, Inc., pp. 2474–2482.
- Powell, W. B., Ryzhov, I. O., 2012. *Optimal learning*. Vol. 841. John Wiley & Sons.
- Priekule, L., Meisel, S., 2017. A bayesian ranking and selection problem with pairwise comparisons. In: *Proceedings of the 49th conference on Winter simulation*. IEEE Computer Society, pp. 2149–2160.
- Shah, N., Balakrishnan, S., Guntuboyina, A., Wainwright, M., 2016. Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In: *33rd International Conference on Machine Learning*. pp. 11–20.
- Shah, N., Wainwright, M., 2016. Simple, robust and optimal ranking from pairwise comparisons. *arXiv preprint arXiv:1512.08949*.
- Suh, C., Tan, V., Zhao, R., 2017. Adversarial top- $k$  ranking. *IEEE Transactions on Information Theory* 63 (4), 2201–2225.
- Thurstone, L., 1927. A law of comparative judgment. *Psychological Review* 34, 273–286.
- Tong, Y. L., 1980. *Probability inequalities in multivariate distributions*. Academic Press.
- Urvoy, T., Clerot, F., Féraud, R., Naamane, S., 2013. Generic exploration and k-armed voting bandits. In: *30th International Conference on Machine Learning*. pp. 91–99.
- Yue, Y., Broder, J., Kleinberg, R., Joachims, T., 2012. The k-armed dueling bandits problem. *Journal of Computer and System Sciences* 78 (5), 1538–1556.
- Yue, Y., Joachims, T., 2011. Beat the mean bandit. In: *28th International Conference on Machine Learning*. pp. 241–248.

## Appendix

Here we give a proof that ranking alternatives by Borda score from pairwise comparisons will correctly reconstruct an underlying latent ranking given certain conditions. For simplicity the proof is given for only binary pairwise sample outcomes, but the proof for unbounded, value-based pairwise samples requires only minor modification.

**Proposition 4.** Given a set of alternatives  $\mathcal{A}$  and some underlying total ordering  $\succeq$  on  $\mathcal{A}$ , suppose that:

- For each pair of alternatives  $(a_i, a_j)$  there is associated a Bernoulli random variable  $X_{i,j}$  and that these random variables are paired such that  $X_{i,j} = 1 - X_{j,i}$ .
- There exists a function  $F : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}$  defined as:

$$F(a_i, a_j) = \mathbb{E}[X_{i,j}]$$

with the following properties:

1. **Strong Stochastic Transitivity** (SST) (Shah et al., 2016) with respect to  $\succeq$ : For alternatives  $a_i, a_j, a_k$ :

$$a_i \succeq a_j \succeq a_k \implies F(a_i, a_k) \geq \max\{F(a_i, a_j), F(a_j, a_k)\}$$

2. **Pairwise Distinguishability**: For any two distinct alternatives  $a_i$  and  $a_j$ , there exists some alternative  $a_k$  such that:

$$F(a_i, a_k) \neq F(a_j, a_k)$$

We define the *Borda Score* ordering  $\geq_B$  on  $\mathcal{A}$  as follows:

$$a_i \geq_B a_j \iff B(a_i) = \sum_{a_k \neq a_i} F(a_i, a_k) \geq \sum_{a_k \neq a_j} F(a_j, a_k) = B(a_j)$$

Then  $\succeq$  and  $\geq_B$  are equivalent, i.e:

$$a_i \succeq a_j \iff a_i \geq_B a_j$$

*Proof.* We begin by showing that:

$$a_i \succeq a_j \implies a_i \geq_B a_j \tag{A.1}$$

To do this, we aim to show that the SST condition implies that all terms in the Borda Score  $B(a_i)$  of  $a_i$  will be at least as large as the corresponding terms in  $B(a_j)$ .

First note that for any alternative  $a_i$ ,  $F(a_i, a_i) = 1 - F(a_i, a_i) = \frac{1}{2}$ . Hence, by the SST condition, for any two alternatives  $a_i$  and  $a_j$ :

$$a_i \succeq a_j \implies F(a_i, a_j) \geq \max\{F(a_i, a_i), F(a_i, a_j)\} \geq \frac{1}{2} \tag{A.2}$$

and so

$$F(a_j, a_i) \leq \frac{1}{2}$$

Now consider the terms in the sums  $B(a_i) = \sum_{a_k \neq a_i} F(a_i, a_k)$  and  $B(a_j) = \sum_{a_k \neq a_j} F(a_j, a_k)$ . For each alternative  $a_k$ , we have three possible cases:

1.  $a_i \succeq a_j \succeq a_k$ :

This case is simple. By the SST condition:

$$F(a_i, a_k) \geq \max\{F(a_i, a_j), F(a_j, a_k)\} \geq F(a_j, a_k)$$

2.  $a_i \succeq a_k \succeq a_j$ : From Equation (A.2) we have that:

$$F(a_i, a_k) \geq \frac{1}{2} \geq F(a_j, a_k)$$

3.  $a_k \succeq a_i \succeq a_j$ : From the SST condition and Equation (A.2), we have:

$$F(a_k, a_j) \geq \max\{F(a_k, a_i), F(a_i, a_j)\} \geq \frac{1}{2}$$

And hence, as  $F(a_k, a_j) = 1 - F(a_j, a_k)$ , we have:

$$F(a_j, a_k) \leq \min\{F(a_i, a_k), F(a_k, a_i)\} \leq F(a_i, a_k)$$

So for all  $a_k$  we have that  $F(a_i, a_k) \geq F(a_j, a_k)$ . Hence  $\sum_{a_k \neq a_i} F(a_i, a_k) \geq \sum_{a_k \neq a_j} F(a_j, a_k)$  and so  $a_i \geq_B a_j$ .

Now we need to prove the converse, i.e:

$$a_i \geq_B a_j \implies a_i \succeq a_j \tag{A.3}$$

Our approach here is slightly different. First, we show that:

$$B(a_i) = B(a_j) \iff a_i = a_j \tag{A.4}$$

If  $a_i = a_j$  it is trivial that  $B(a_i) = B(a_j)$ . So now let us suppose that  $a_i \neq a_j$ . As  $\succeq$  is a total ordering, exactly one of either  $a_i \succeq a_j$  or  $a_j \succeq a_i$  must be true, so let us assume without loss of generality that  $a_i \succeq a_j$ . Now, as  $a_i \neq a_j$ , and alternatives are *Pairwise Distinguishable* under  $F$ , there exists some  $a_{k'}$  such that  $F(a_i, a_{k'}) \neq F(a_j, a_{k'})$ . As we have shown above,  $\forall a_k, F(a_i, a_k) \geq F(a_j, a_k)$ , and thus, we must have that for  $a_{k'}$ ,  $F(a_i, a_{k'}) > F(a_j, a_{k'})$  and so  $B(a_i) > B(a_j)$ . By contraposition, this proves Equation (A.4).

As  $\succeq$  is reflexive, Equation (A.4) implies that for  $a_i = a_j$ :

$$a_i \geq_B a_j \implies a_i \succeq a_j$$

So it only remains to show this for  $a_i \neq a_j$ . But, in proving Equation (A.4), we have already shown that if  $a_i \neq a_j$ :

$$a_i \succeq a_j \implies B(a_i) > B(a_j) \tag{A.5}$$

Thus the contrapositive of Equation (A.5) is also true, specifically:

$$B(a_j) \geq B(a_i) \implies !(a_i \succeq a_j) \implies a_j \succeq a_i$$

With the last implication justified by the totality of the ordering  $\succeq$ .  $\square$

So, if we have Strong Stochastic Transitivity (SST), and Pairwise Distinguishability (PD), given sufficiently accurate estimates for pairwise means, the Borda Score ranking for alternatives will be consistent with an underlying latent ranking, if one exists. We should consider how reasonable these two conditions are.

Firstly, remember that the only method for gaining information about alternatives is through pairwise measurements. Therefore, if PD does not hold, and there are some alternatives for which all pairwise means are identical, then it is impossible for us to gain any information that would allow us to distinguish one such alternative from another. Thus, PD is a requirement for any pairwise ranking problem where we might wish to create a total ordering of alternatives.

The SST condition is a common condition for modelling pairwise ranking and selection problems, being either explicitly or implicitly assumed in many of the current methods discussed in Section 3. It also holds for the most common latent fitness models (BTL and Thurstone), as well as generally being consistent with observed real-world pairwise preference data Shah et al. (2016). Note that the POCBAM and PKG methods proposed in this paper are highly general, and do not require that the SST condition holds for pairwise outcomes, it is only required to guarantee equivalence between the Borda Score ordering and an underlying latent ordering, if one exists.

**Proof of Inequality 4.1.3** Let  $Y \sim \mathcal{N}(M, \Sigma)$  be the joint distribution of the alternative score estimates of the current top- $\kappa$  alternatives ( $M_i = \tilde{\mathcal{S}}_i$ ), and define  $Y^* \sim \mathcal{N}(M, \Sigma^*)$  to be the multivariate Gaussian distribution with equal mean to  $Y$ , and covariance matrix defined by:

$$\Sigma_{i,j}^* = \begin{cases} \Sigma_{i,j}, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Then  $\Sigma_{i,j}^* \geq \Sigma_{i,j}, \forall i, j$ . Hence by Theorem 2.1.1, Corollary 1 of Tong (1980):

$$\mathbb{P}_Y \left[ \bigcap_{i=1}^{\kappa} \{\tilde{\mathcal{S}}_i \geq a_i\} \right] \leq \mathbb{P}_{Y^*} \left[ \bigcap_{i=1}^{\kappa} \{\tilde{\mathcal{S}}_i \geq a_i\} \right] = \prod_{i=1}^{\kappa} \mathbb{P}(\tilde{\mathcal{S}}_i \geq a_i)$$

Holds for any  $(a_1, \dots, a_{\kappa}) \in \mathbb{R}^{\kappa}$ . The proof of the second inequality for alternatives outside the current top- $\kappa$  follows from Theorem 2.1.1 of Tong (1980) by an almost identical argument.